

Osteoarthritis and Cartilage



Total knee replacement; minimal clinically important differences and responders



A. Escobar ^{†,*}, L. García Pérez [‡], C. Herrera-Espiñeira [§], F. Aizpuru ^{||}, C. Sarasqueta [¶],
M. Gonzalez Sáenz de Tejada [†], J.M. Quintana [†], A. Bilbao [†]

[†] Research Unit, Hospital Universitario Basurto, Avda. Montevideo 18, 48013 Bilbao, Spain

[‡] Planning and Evaluation Service, Canary Islands Health Service, C/Pérez de Rozas, n°5, 4th Floor, 38004 Santa Cruz de Tenerife, Spain

[§] Hospital Virgen de las Nieves, Avda. de las Fuerzas Armadas, 2, Granada, Spain

^{||} Research Unit, Hospital Universitario Araba, C/Jose Atxotegi s/n, 01006 Vitoria-Gasteiz, Spain

[¶] Research Unit, Hospital Universitario Donostia, Paseo Dr. Beguiristain s/n, 20014 San Sebastián, Donostia-San Sebastian, Spain

[#] Research Unit, Hospital of Galdakao-Usansolo, B° Labeaga s/n, Spain

ARTICLE INFO

Article history:

Received 22 May 2013

Accepted 24 September 2013

Keywords:

Total knee replacement

MCID

Responders

SUMMARY

Objective: To provide new data on minimally clinical important difference (MCID) and percentages of responders on pain and functional dimensions of Western Ontario and McMaster Osteoarthritis Index (WOMAC) in patients who have undergone total knee replacement (TKR).

Methods: 1-year prospective multicentre study with two different cohorts. Consecutive patients on the waiting list were recruited. There were 415 and 497 patients included. Pain and function were collected by the reverse scoring option of the WOMAC (0–100, worst to best). Transition items (five point scale) were collected at 1-year and MCID was calculated through mean change in patients somewhat better, Receiver Operating Characteristic (ROC) and two other questions about satisfaction. Analysis was performed in the whole sample and by tertiles of baseline severity. Likewise were calculated the percentages of patients who attained cut-off values.

Results: Global MCID for pain were about 30 in both cohorts and 32 for. By ROC these values were about 20 and 24 respectively. According to the other two transitional questions these values were for pain 27 and 20 for function. By tertiles the worst the baseline score the higher the cut-off values. Percentage of responders does not change when comparing responders to the global MCID with their own tertile MCID and were about 61% for pain and 50% for function.

Conclusion: Due to the wide variations, MCID estimates should be calculated and used according to the baseline severity score.

© 2013 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.

Introduction

When considering the expected benefits of total knee replacement (TKR) we are, basically thinking in terms of pain relief, functional recovery and health-related quality of life. Overall, literature tells us that TKR is an effective intervention in terms of patient reported outcomes (PROs)^{1,2}.

But is this intervention really as effective as we think? A recent study has reported about 44% of patients with persistent pain 3–4 years after the TKR, with even 15% reported as severe-extreme³. One systematic review⁴ stated that after TKR there was unfavourable long-term pain outcome in 10–34%. Furthermore, 10–30% of patients report poor outcomes after intervention in terms of pain at 1–7 years^{3,5}. Most of the outcomes studied are focused on pain

* Address correspondence and reprint requests to: A. Escobar, Unidad de Investigación, Hospital Universitario Basurto, Avda. Montevideo 18, 48013 Bilbao, Bizkaia, Spain. Tel: 34-944006309; Fax: 34-944006306.

E-mail addresses: antonio.escobarmartinez@osakidetza.net (A. Escobar), lidia.garciaperez@sescs.es (L. García Pérez), carmen.herrera.sspa@juntadeandalucia.es (C. Herrera-Espiñeira), felipeesteban.aizpurubaranandiaran@osakidetza.net (F. Aizpuru), cristina.sarasquetaeizaguirre@osakidetza.net (C. Sarasqueta), marta.gonzalezsaenzdetejada@osakidetza.net (M. Gonzalez Sáenz de Tejada), josemaria.quintanalopez@osakidetza.net (J.M. Quintana), Amaia.bilbaogonzalez@osakidetza.net (A. Bilbao).

^a Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Kronikgune.

rather than on functionality and this may be because pain is easier to improve than function due to the mean age of these patients. In addition pain is the main complaint for patients electing to undergo TKR. Functionality related to daily living activities also improves but this may be closely related to the decrease in pain.

PRO questionnaires are one way to know how many patients can be considered as responders after TKR. They are increasingly used in clinical practice, but to add some valid information to clinicians they should be easily interpretable. There are several ways to facilitate interpretation of these questionnaires. One of the most used is the Minimally Clinically Important Difference (MCID)⁶. More recently, the concept of Patient Acceptable Symptom State (PASS)^{7,8} defined as the value beyond which patients consider themselves well and the Minimal Clinically Important Improvement (MCII)⁹ defined as the smallest change in measurement that signifies an important improvement in a patient's symptom, have appeared. These concepts are complementary; all are based on external anchoring questions but while MCID and MCII measure improvement (feeling better), the PASS measures a satisfactory final state (feeling good).

Regarding MCID it is widely recognized that there are no universal MCID values for a given PRO, but it may vary by patient's characteristics or different clinical context. Therefore, the MCID of an instrument cannot be trusted if determined in only one study or by one method².

In TKR, there are few studies^{10–13} which provide data on the MCID. Three of them^{10,11,13} provide data at 6 months or 2 years. In the remaining¹² the MCID cut-off points were studied at 1 year after surgery. On the other hand it has been argued that MCID based on a sample global value may result in a significant proportion of individuals being misclassified as having no benefit and, therefore, MCIDs should be reported taking into account the baseline score¹⁴.

The main goal of this study was to provide new data on MCID and responders at 1 year in patients who have undergone TKR, measured by pain and functional dimensions of Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) which could facilitate the interpretation of WOMAC changes.

Methods

We have included two cohorts of patients waiting for TKR. The first study took place in six public hospitals belonging to the Basque Health Service-Osakidetza, a local government agency in the Basque Country. The second cohort is also a 1-year prospective study that took place in 15 hospitals; three in Andalusia, three in the Canary Islands and nine in the Basque Country (Spain). Consecutive patients placed on the waiting list to undergo primary TKR for osteoarthritis between September 2003 and September 2004 and between March 2005 and December 2006 and managed in any of the hospitals were eligible for the study. In both studies, patients with psychiatric diseases were excluded because these could prevent them from filling out the questionnaires of the study. We collected data from medical records and directly from patients. We sent to the patients questionnaires at baseline and 12 months post-surgery. All patients received a letter informing them about the study and asking for their voluntary participation. The Institutional Review Boards of the Hospitals approved both studies. The data used in this study comprise a subset of patients who have completed preoperative and postoperative health related quality of life questionnaires and all the transition questions.

We used the WOMAC that is a disease-specific, self-administered questionnaire¹⁵. It has a multidimensional scale made up of 24 items grouped into three dimensions: pain (five items), stiffness (two items), and physical function (17 items). We have studied pain and function dimensions through the Likert version with five response levels, representing different degrees of intensity: none

(0), mild (1), moderate (2), severe (3) or extreme (4). The final scores were determined by adding the corresponding items for each dimension, and standardizing to a range of values from 0 to 100. According to recent recommendations¹⁶ we have used the reverse option, from 0 (worst) to 100 (best). The WOMAC has been translated and validated into Spanish^{17,18}.

Statistical analysis

Descriptive statistics included frequencies and percentages for categorical variables and means and standard deviations (SDs) for continuous variables. We compared between both cohorts the sociodemographic data and WOMAC domains at baseline and 1 year after the intervention. Chi-square test was used for comparing categorical variables, whereas Student's *t* test or the nonparametric Wilcoxon test was used for continuous variables.

We used different statistical methods to calculate the cut-off values for MCID which has been defined⁶ as the smallest difference between the scores in a questionnaire that the patient perceives to be beneficial. All patients had to answer two raw transition items (RTI), about their improvement or deterioration, one about pain and another about function 1 year after TKR (Compared to before surgery, how would you rate pain (functional limitation) in the same knee?). The five responses were "a great deal better", "somewhat better", "equal", "somewhat worse" and "a great deal worse". The MCID was estimated for both domains by the mean change score for those patients who, declared changes "somewhat better" in the RTI taking the difference between the score at 1 year and at the baseline¹⁹. Second, we have used the Receiver Operating Characteristics (ROC) curve approach, considering the dichotomized RTI (a great deal better and somewhat better vs equal, somewhat worse and a great deal worse) as the dependent variable, and the change score for each dimension as independent. As optimal cut-off value of each dimension, the one which maximized the sum of sensitivity and specificity was considered. We draw 500 bootstrap samples²⁰, calculated their respective ROC curves and derived the 95% confidence interval (CI).

To assess the usefulness of RTI in establishing the MCID, we have evaluated their validity and reliability¹². Validity through the association between RTI and the change score in pain and function, by means of partial correlation coefficients, controlling for baseline score. We hypothesized that correlation should be higher than 0.5²¹. We evaluated the correlation among RTI and pre and post-scores by Spearman's correlation coefficient. Reliability (internal consistency) was studied by Cronbach's alpha of the two RTI, pain and function items combined. According to recommendations, we have estimated the standard error of measurement (SEM), which represents the amount of error associated with an individual subject assessment by the formula $SEM = SD [(1 - R)^{1/2}]$ where SD is the baseline standard deviation and *R* is the Cronbach's alpha reliability coefficient in our samples. At the individual level, SEM should be smaller than MCID/4 to distinguish MCID from measurement error²².

As recommended²³ we have also calculated MCID through two different independent anchor questions. First: Was the surgery worthwhile? The answers were: definitely yes, probably yes, probably not and definitely not. The MCID was calculated in patients who scored "probably yes". The second question was: What is your global level of satisfaction with surgical management? The answers were, very satisfied, somewhat satisfied, somewhat dissatisfied and very dissatisfied. We considered "somewhat satisfied" in calculating MCID. We have chosen these answers because are the closest to "somewhat better" in the RTI about pain and function dimensions.

For each cut-off value, the percentage of patients exceeding the MCID (responders) with the 95% CI was estimated. Analyses were performed globally and according to baseline severity tertiles.

We have studied satisfaction of patients through the next question: “If you had to be the rest of your life with the symptoms you have now, how would you feel? Response categories were very satisfied, somewhat satisfied, somewhat dissatisfied and very dissatisfied. For analysis we grouped the first two into satisfied patients⁸.

We considered statistically significant $P < 0.05$. All statistical analyses were performed using SAS for Windows version 9.2 (SAS Institute, Inc., Cary, NC), and Confidence Interval Analysis v.2.2.

Results

Samples description

There were 415 and 497 patients in the first and second cohorts respectively. In both groups, about 70% were females, the mean age was 71 years old and the mean Body Mass Index (BMI) was 30. As can be seen in Table I both in total sample and sample data divided by baseline severity tertiles, there were no statistical differences between cohorts, except in the scores at 1 year in the functional dimension. As it was expected, there were large improvements, both in pain and function, about 34 and 32 points, respectively. In both groups from the worst to best tertile, gains in pain were about 51, 33 and 18 points respectively, and about 46, 33 and 17 points in the function dimension. This is, the worse the baseline status the higher the gains.

In comparing baseline pain, function, age, BMI and gender, there were no differences between patients included and not included in the first cohort. In the second, non-included patients scored five points higher in pain and function and, there were 6% more females (data not shown).

RTI

The partial correlation coefficients between RTI-change scores in pain were 0.53 (first cohort) and 0.62 (second cohort), while for function they were 0.42 and 0.60, respectively. The correlation between RTI-baseline pain was 0.03 and -0.05 in the first and second cohort, while with the 1-year score it was 0.62 and 0.47. As regards function, these values were 0.03 and -0.08 for baseline scores and 0.60 and 0.48 for 1 year, respectively. Reliability values for the combined RTI were 0.70 and 0.80 in the first and second cohort, respectively.

MCID for pain

Table II shows data on the SEM and MCID in the pain dimension with their 95% CI along with the percentage of patients who were above those values. The SEMs were 8.6 and 8.3 in the first and second cohort. Regarding global MCID is about 29 points in both cohorts. The values according to the baseline tertiles of severity were similar in both cohorts, around 45, 28 and 15 points in the worst, medium and best tertiles, respectively. The global value obtained by ROC analysis was about 22 points. There was a major difference in the worst tertile, being 30 points in the first cohort and 40 in the second one, while the other two tertiles were quite similar, around 23 in medium and 10 points in the best tertile. The MCID according to the two additional transitional questions, were around 27 points, similar and close to the global MCID.

MCID for function

In Table III there is the same data as previously shown but on the function dimension. The SEM values were 5.1 in the first cohort and 4.6 in the second. Global MCID was about 32 points. According to

Table I
Sample characteristics by cohort

	First cohort (n = 415)	Second cohort (n = 497)	P-value
Female, n (%)	296 (71.3)	345 (69.4)	0.6
Age	Mean (SD)	Mean (SD)	
	71.4 (6.7)	71.4 (7.0)	1
BMI*	30.1 (4.4)	30.3 (4.8)	0.5
WOMAC† scores			
Presurgery pain	44.2 (17.3)	44.6 (18.7)	0.7
Worst tertile	24.4 (10.2)	24.3 (10.4)	1
Medium tertile	45.4 (4.3)	44.9 (4.1)	0.4
Best tertile	63.1 (8.5)	64.3 (9.6)	0.2
Pain at 1 year	78.1 (16.4)	78.7 (19.6)	0.6
Worst tertile	74.4 (20.0)	76.1 (21.3)	0.5
Medium tertile	78.4 (14.2)	77.2 (18.9)	0.5
Best tertile	81.5 (13.8)	82.7 (18.0)	0.5
Presurgery function	38.3 (17.3)	38.8 (17.6)	0.7
Worst tertile	19.6 (7.7)/141	19.4 (8.4)/164	0.8
Medium tertile	38.1 (4.7)/134	39.0 (4.2)/173	0.08
Best tertile	57.6 (8.6)/140	58.5 (9.9)/160	0.4
Function at 1 year	69.3 (18.9)	72.0 (20.8)	0.04
Worst tertile	65.5 (22.2)	65.8 (23.3)	0.9
Medium tertile	69.2 (17.3)	73.4 (18.9)	0.04
Best tertile	73.2 (16.1)	76.6 (18.7)	0.09

Tertiles of pain: first cohort (second cohort): worst: ≤ 37.5 (≤ 35); medium: 38–50 (35.5–50); best: > 50 (> 50)

Tertiles of function first cohort (second cohort): worst ≤ 29.4 (≤ 31.7); medium: 29.5–45.6 (31.8–45.6); best: > 45.6 (> 45.6)

Sample size in each tertile first cohort (second cohort): worst: 134 (169); medium: 147 (156); best: 134 (172)

* Body mass index

† WOMAC range from 0 to 100; worst to best

the baseline tertiles of severity, both groups showed similar values, which from worst to best were around 45, 33 and 17. The cut-off values established by ROC were also similar in both groups. The global value was around 24 points, and from worst to best tertile were 40, 27 and 15 points. Data about the cut-off points in the two other transitional questions had a mean value of 22 points (range: 18.8–25.8).

Responders

Table IV shows data on percentage of responders and their 95% CI according to the values of the global MCID, by tertile of baseline scores and satisfied patients. According to the global MCID, the percentage of responders in pain was similar between cohorts, 61.9% and 63.4%. These percentages increased as the baseline severity increased (more responders in the worst tertile) with similar percentages in both cohorts in all tertiles, varying from 31.6% to 87%. According to the MCID by tertiles, the total percentage of patients considered as responders was 60.9% and 69.6% in each cohort respectively. According to the tertiles these percentages ranged from 55.9% to 66.9% in the first and from 66.0% to 71.6% in the second cohort. When considering the function dimension, the percentages of responders to the global MCID were 48.7% and 49.8% in the first and second cohort. As before, these percentages were higher in the worst tertile ranging from 20.6% to 74.1% in the first cohort and from 18.5% to 72.4% in the second one. If we look at the responders according to their own MCID by tertile, patients classified as responders were 50.4% and 53.0%. These percentages ranged from 46.3% to 55.2% and from 51.6% to 55.2% in the first and second group respectively.

Satisfaction

Satisfaction with their “current” symptomatology was attained by 78.5% and 78.9% of patients in each cohort. By tertiles of pain

Table II
MCID data for the WOMAC pain domain

	First cohort (n = 415)		Second cohort (n = 497)	
	Cut-off value (95% CI)	Patients %§	Cut-off value (95% CI)	Patients %§
SEM*	8.6		8.3	
MCID† global‡	29.9 (27.1–32.6) n = 209	61.9	28.1 (25.1–31.0) n = 207	63.4
MCID† tertiles‡				
Worst	46.4 (40.6–52.2) n = 59	60.9	44.5 (39.9–49.2) n = 67	69.6
Medium	29.9 (26.7–33.1) n = 75		27.1 (23.4–30.7) n = 71	
Best	16.8 (13.6–20.0) n = 75		13.1 (8.8–17.4) n = 69	
ROC global	20.5 (20.2–20.9)	70.2	23.5 (23.1–23.8)	72.4
Worst	29.8 (29.2–30.5)	84.6	40.7 (40.2–41.1)	71.6
Medium	21.1 (20.7–21.5)	78.4	25.0 (24.7–25.3)	66.0
Best	11.7 (11.2–12.2)	71.3	9.0 (7.8–10.2)	75.0
Was the surgery worthwhile?				
- Patient's answer: Probably YES, n (%)	96 (23.1)		103 (20.7)	
MCID†	28.0 (23.6–32.4)	61.3	25.6 (21.6–29.7)	64.0
What is your global level of satisfaction with surgical management?				
- Patient's answer: Somewhat satisfied, n (%)	115 (27.7)		138 (27.8)	
MCID†	25.2 (21.9–28.5)	61.3	27.5 (23.7–31.4)	63.4

Tertiles of pain: first cohort (second cohort): worst: ≤ 37.5 (≤ 35); medium: 38–50 (35.5–50); best: > 50 (> 50).

ROC: calculated as the point that maximized the sum of sensitivity and specificity.

* SEM: standard error of measurement.

† MCID: Minimal clinically important difference.

‡ Calculated as mean change in those patients who were “somewhat better”.

§ Percentage of patients exceeding the cut-off value.

|| Sample size in the “somewhat better category”.

severity, percentages varied from 77.8% to 79.2% in the first cohort, and from 75.1% to 84.9% in the second. By tertiles of function, they ranged from 75.6% to 81.6% and from 74.8% to 82.0% in each cohort respectively.

Discussion

In this study of two cohorts we have used several ways to establish cut-off values for patients to be considered as responders according to MCID. The fact of using two cohorts could be seen as a way of validating the values obtained. This study provides more information on the improvement scores which can be used as a reference when considering WOMAC after TKR.

We may consider our RTI correlation values as the usual reported in the literature regarding their reliability and validity. All RTI-change score partial correlations were above 0.5 except that of function in the first cohort, which was in the vicinity of this threshold value. As it was expected, we have obtained a high positive correlation with the 1-year score and a near zero correlation with baseline scores in both dimensions which implies that patients do not remember properly their baseline state²⁴. Reliability measured by Cronbach's alpha showed values considered as acceptable or good. In both cases our data were slightly lower than others published¹².

There are some published MCID values for WOMAC based on absolute change^{10–12}. Values are different and this could be interpreted in at least three ways. First, the timeframe in which values

Table III
MCID data for the WOMAC function domain

	First cohort (n = 415)		Second cohort (n = 497)	
	Cut-off value (95% CI)	Patients %	Cut-off value (95% CI)	Patients %
SEM*	5.1		4.6	
MCID† global‡	31.1 (28.3–33.9) n = 221¶	48.7	33.5 (30.9–36.0) n = 237¶	49.8
MCID† tertiles‡				
Worst	44.1 (39.4–48.9) n = 76¶	50.4	45.6 (40.9–50.3) n = 81¶	53.0
Medium	32.8 (29.1–36.5) n = 70¶		33.9 (30.6–37.1) n = 82¶	
Best	16.3 (12.8–19.9) n = 75¶		19.7 (16.7–22.7) n = 74¶	
ROC§ global	24.2 (23.6–24.7)	61.2	23.0 (22.7–23.2)	69.7
Worst	43.3 (42.6–44.0)	58.0	39.7 (39.1–40.4)	61.3
Medium	26.8 (26.3–27.3)	60.3	27.1 (26.7–27.5)	69.8
Best	15.3 (14.9–15.8)	52.5	14.7 (14.3–15.0)	65.6
Was the surgery worthwhile?				
- Patient's answer: Probably YES, n (%)	96 (23.1)		103 (20.7)	
MCID†	20.6 (16.4–24.8)	84.6	22.5 (19.0–26.1)	68.9
What is your global level of satisfaction with surgical management?				
- Patient's answer: Somewhat satisfied, n (%)	115 (27.7)		138 (27.8)	
MCID†	18.8 (15.7–21.9)	86.5	25.8 (22.4–29.2)	63.0

Tertiles of function first cohort (second cohort): worst ≤ 29.4 (≤ 31.7); medium: 29.5–45.6 (31.8–45.6); best: > 45.6 (> 45.6).

* SEM: standard error of measurement.

† MCID: Minimal clinically important difference.

‡ Calculated as mean change in those patients who were “somewhat better”.

§ Receiver Operating Characteristic: calculated as the point that maximized the sum of sensitivity and specificity.

|| Percentage of patients exceeding the cut-off value.

¶ Sample size in the “somewhat better category”.

Table IV

WOMAC pain and function dimensions: percentage of patients attaining global MCID, their own MCID by tertile and satisfied patients

	Baseline tertile, %			Total, % (95% CI)
	Worst	Medium	Best	
Pain dimension				
First* cohort (n = 415)				
MCID* global	84.6	66.9	31.6	61.9 (57.1–66.7)
MCID by tertile	57.4	66.9	55.9	60.9 (56.1–65.8)
Satisfied†	77.8	78.4	79.2	78.5 (74.4–82.6)
Second* cohort (n = 497)				
MCID global	87.0	65.4	38.4	63.4 (59.1–67.7)
MCID by tertile	71.6	66.0	70.9	69.6 (65.5–73.8)
Satisfied†	75.1	76.3	84.9	78.9 (75.2–82.6)
Function dimension				
First* cohort (n = 415)				
MCID global	74.1	49.3	20.6	48.7 (43.7–53.6)
MCID by tertile	55.2	46.3	47.5	50.4 (45.4–55.3)
Satisfied†	78.9	75.6	81.6	78.5 (74.4–82.6)
Second* cohort (n = 492)				
MCID global	72.4	57.0	18.5	49.8 (45.3–54.3)
MCID by tertile	55.2	52.3	51.6	53.0 (48.5–57.6)
Satisfied†	74.8	82.0	79.9	78.9 (75.2–82.6)

* MCID: Minimal Clinically Important Difference.

† Question was: "If you had to be the rest of your life with the symptoms you have now, how would you feel?" Responses categories: very satisfied, somewhat satisfied, somewhat dissatisfied and very dissatisfied.

are measured. Possibly it is not the same to measure outcomes in patients with TKR at 6 months or at 2 years¹⁰. Second, these are values that strongly rely on the kind of population included in the study; threshold of improvement should be different in patients with TKR than in those with medical treatment or patients included in a rehabilitation programme. The fact that those TKR patients who are located in the best pain tertile have much lower MCID values than their sample global MCID and close to those with medical treatment, about 15.5²⁵ could serve as a reference. Third, values rely on how they are calculated, as a mean change in one category in the RTI, as our work or as the difference in mean change between adjacent categories. These facts should be taken into account when comparing MCID cut-off points. Our work presents data at 1 year for TKR and calculated through RTI taking patients classified as "somewhat better" to establish changes.

Our data show an important improvement in the pain and functional dimensions at 1 year post-surgical management. As in others studies^{1,26} the worse the baseline score the higher the improvement in both dimensions. A recent review⁴ reported that the proportion of patients with long-term pain outcome after TKR was high and in those studies with the best quality this proportion was about 20%. In this paper, when using WOMAC pain dimension to measure unfavourable pain outcome these percentages varied according to the chosen criteria. For instance, using as a cut-off value 10/100, there was an 18.5% of patients who did not improve; our data (not shown) are slightly lower with 14.3% and 12.9% in each group. When using MCID of 22.6 points calculated with one transition question for both dimensions¹³ this percentage was 33.1% while in our series with two transitions questions these cut-off values given percentages of 29.8% and 27.6% (data not shown) in the first and second cohort respectively. According to our results, there are about 38.1% (first cohort) or 36.6% (second) of patients who have not reached the global MCID for pain.

It is well known^{9,14} that MCID can change according to baseline severity. Therefore, it is considered appropriate to establish cut-off points according to them in order to calculate the percentage of patients who can be considered as responders within their own baseline tertile, rather than to a sample MCID. Our data show different cut-off points in each tertile, the worse the baseline status

the higher the needed score, and different percentages of responders according to their own tertile. Cut-off points were very different in each tertile but percentages of responders were similar according to the global MCID or to the MCID by tertile. Therefore, although MCID should be calculated by tertiles, this fact does not change the overall percentage of patients considered as responders. Therefore, with our data there is no important effect of the baseline score. For instance, in our second cohort, patients who are considered as responders to their own tertile MCID and not to global MCID (n = 57) were all located in the best baseline situation, and *vice versa*, there were 26 patients who were responders to global MCID and not to their own tertile, all of them were in the worst baseline situation. So the misclassification of patients based on global MCID plays a role in both directions.

Although we agree with Davis *et al.*¹⁴ in the existence of a third group of patients who would be classified as non-responders due to the ceiling (or floor) effect of a questionnaire and that there are another variables apart from pain and functional ability, which could influence the decision to perform TKR^{27,28}, our percentages of patients who could not reach the global sample MCID was much lower than those reported by them. In the pain dimension we had a 4.3% and 6.8% in the first and second cohort of patients who could not reach MCID. In the previous report there was 28.8%. Clearly, all these patients are in the lower pre-surgery pain tertile, which is the best situation. When groups have this or similar percentages of patients located in the best baseline situation, analysis by tertiles can have higher importance given that those are the patients who cannot attain global MCID but more easily reach the MCID of their baseline tertile. Although there are conflicting results in the literature^{29,30}, these differences may be due to the time the baseline score was collected, possibly the time on the waiting list could influence it, or it can be affected when others factors apart from pain or functional disability have a major influence on decisions to surgery.

As in other study¹², we have considered helpful to use another external criterion of change, such as those presented in Tables II and III. In the case of pain, the thresholds of these questions are slightly lower than MCID with their own RTI, as in the cited study. With these data, we could state that the global MCID for pain in patients with TKR at 1 year would be around 30 points. This figure represents a 30% improvement over the baseline score which agrees with other consensus statements on pain³¹ and a validation study of this statement using an objective functional external anchor³² in pain or function carried out in patients with lumbar disorders. It was stated that a change of at least 30% was considered as moderately improvement. We may also say that this value is similar to that reported by Chesworth *et al.*¹².

As before, cut-off values calculated through the ROC analysis varies depending on the baseline score being the worst patients who need the greater improvement to attain them. ROC analysis consistently presents more conservative values for MCID. However, these values are based on a small number of non-improved patients indicating poor reliability.

Considering satisfaction with symptomatology, the percentage of patients who considered themselves satisfied is more stable along tertiles of baseline severity and always higher than responders to the MCID by tertile. This question has been previously considered as a PASS estimate⁸ and shows that the percentage of responders can be considered independently of baseline scores. Our percentage of satisfied patients, around 80%, is similar to others published^{33–35}.

Regarding physical functioning, our global MCID was about 32 points and it is close to another reported value of 33 points¹² and our 23–24 points calculated by ROC is also similar to 22.8 points reported by Terwee *et al.*¹¹. In addition, results of both of our cohorts are very similar. In this case our estimated values are higher than

those calculated with one transition question¹⁰ at 6 months (19.1) or at 2 years (20.8). As in the pain dimension, cut-off points by tertiles were higher in the worst baseline status in all three methods. However, in this dimension, MCID calculated by the two other anchor questions, were lower and closest to the ROC values than to the global MCID. All comments on responders vs non-responders and analysis by tertiles performed on pain are applicable to the function dimension.

If we consider the percentage of global MCID responder and those of responders according to their own tertile, we can see that both are similar with a slight increase if we consider data divided by tertiles.

There are some possible limitations in the study. It is well known that the use of retrospective RTI is subject to recall bias, so the validity of these retrospective items could be threatened. Patients non-included in the study could bias the results, but our data do not seem to appreciate differences between participants and excluded patients. ROC values are based on a small number of non-improved patients, so the result can be unreliable. Finally, our pain SEM values were high and therefore don't seem to be adequate to use in individual patients, although our values of MCID/4 are close to the SEM.

In conclusion, data of both groups are similar; the global MCID can be considered around 30 points and 32 for pain and functional dimensions. Data calculated by ROC analysis are always lower and around 22 points for pain and 23 for function dimension. The worse the baseline status the higher the cut-off point obtained by any method. Although cut-off points should be calculated by baseline tertiles of severity, the percentage of responders is similar if we calculate it by means of the global sample MCID as if we do it by means of the MCID by tertiles.

Author contributions

All authors have contributed to each of three activities:

- (1) the conception and design of the study, or acquisition of data, or analysis and interpretation of data,
- (2) drafting the article or revising it critically for important intellectual content, and
- (3) final approval of the final version, and will take public responsibility for the content of the paper. The content has not been published, nor is it being considered elsewhere.

Conflict of interest

No possible conflicts of interest (e.g., funding sources for consultations or studies of products) exist in this study.

Role of funding sources

Both public sources have only been funders of the research projects. The study sponsors have not been involved in any process of the manuscript.

Acknowledgements

We thank the staff members of the different departments, research units and medical records sections of the participating Hospitals, as well as the patients for their collaboration.

This study was supported by grants from the Fondo de Investigación Sanitaria: First cohort: FIS01/184, Second cohort: FIS (PI 04/0938, PI 04/0542, and PI 04/2577) and from the Basque Country Health Department (200411012).

References

1. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J Bone Joint Surg Am* 2004;86-A(963).
2. Collins NJ, Roos EM. Patient-reported outcomes for total hip and knee arthroplasty: commonly used instruments and attributes of a "good" measure. *Clin Geriatr Med* 2012;28:367.
3. Wylde V, Hewlett S, Learmonth ID, Dieppe P. Persistent pain after joint replacement: prevalence, sensory qualities, and postoperative determinants. *Pain* 2011;152:566.
4. Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open* 2012;2:e000435.
5. Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. *J Rheumatol* 2000;27:1745.
6. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertain the minimal clinically important difference. *Control Clin Trials* 1989;10:407.
7. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheum Dis* 2005;64:34.
8. Escobar A, Gonzalez M, Quintana JM, Vrotsou K, Bilbao A, Herrera-Espineira, et al. Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement. Identification of cut-off values. *Osteoarthritis Cartilage* 2012;20:87.
9. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis* 2005;64:29.
10. Escobar A, Quintana JM, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 2007;15:273.
11. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010;63:524.
12. Chesworth BM, Mahomed NN, Bourne RB, Davis AM. Willingness to go through surgery again validated the WOMAC clinically important difference from THR/TKR surgery. *J Clin Epidemiol* 2008;61:907.
13. Quintana JM, Escobar A, Arostegui I, Bilbao A, Azkarate J, Goenaga JI, et al. Health-related quality of life and appropriateness of knee or hip joint replacement. *Arch Intern Med* 2006;166:220.
14. Davis AM, Perruccio AV, Lohmander LS. Minimally clinically important improvement: all non-responders are not really non-responders an illustration from total knee replacement. *Osteoarthritis Cartilage* 2012;20:364.
15. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1983;15:1988.
16. Singh J, Sloan JA, Johanson NA. Challenges with health-related quality of life assessment in arthroplasty patients: problems and solutions. *J Am Acad Orthop Surg* 2010;18:72.

17. Batlle-Gualda E, Esteve-Vives J, Piera MC, Hargreaves R, Cutts J. Traducción y adaptación al español del cuestionario WOMAC específico para artrosis de rodilla y cadera. *Rev Esp Reumatol* 1999;38.
18. Escobar A, Quintana JM, Bilbao A, Azkarate J, Guenaga JL. Validation of the Spanish version of the WOMAC questionnaire for patients with hip or knee osteoarthritis. Western Ontario and McMaster Universities Osteoarthritis Index. *Clin Rheumatol* 2002 Nov.;466:1921.
19. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD* 2005;2:63.
20. Efron B, Tibshirani R. An Introduction to Bootstrap. Boca Raton: CRC Press LLC; 1994.
21. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. *J Clin Epidemiol* 2002;55:900.
22. Terwee CB, Roorda LD, Knol DL, De Boer MR, de Vet HC. Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol* 2009;62:1062.
23. Revicki DA, Erickson PA, Sloan JA, Dueck A, Guess H, Santanello NC. Interpreting and reporting results based on patient-reported outcomes. *Value Health* 2007;10(Suppl 2): S116–24.
24. Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res* 2005;14:285.
25. Hmamouchi I, Allali F, Tahiri L, Khazzani H, Mansouri LE, Ali Ou AS, et al. Clinically important improvement in the WOMAC and predictor factors for response to non-specific non-steroidal anti-inflammatory drugs in osteoarthritic patients: a prospective study. *BMC Res Notes* 2012;5:58.
26. Nunez M, Nunez E, del Val JL, Ortega R, Segur JM, Hernandez MV, et al. Health-related quality of life in patients with osteoarthritis after total knee replacement: factors influencing outcomes at 36 months of follow-up. *Osteoarthritis Cartilage* 2007;15:1001.
27. Dieppe P, Lim K, Lohmander S. Who should have knee joint replacement surgery for osteoarthritis? *Int J Rheum Dis* 2011;14:175.
28. Escobar A, Quintana JM, Arostegui I, Azkarate J, Guenaga JL, Arenaza JC, et al. Development of explicit criteria for total knee replacement. *Int J Technol Assess Health Care* 2003;57.
29. Hirvonen J, Tuominen U, Seitsalo S, Lehto M, Paavolainen P, Hietaniemi K, et al. The effect of waiting time on health-related quality of life, pain, and physical function in patients awaiting primary total hip replacement: a randomized controlled trial. *Value Health* 2009;12:942.
30. Desmeules F, Dionne CE, Belzile EL, Bourbonnais R, Fremont P. The impacts of pre-surgery wait for total knee replacement on pain, function and health-related quality of life six months after surgery. *J Eval Clin Pract* 2012;18:111.
31. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 2008;9:105.
32. Gatchel RJ, Mayer TG, Choi Y, Chou R. Validation of a consensus-based minimal clinically important difference (MCID) threshold using an objective functional external anchor. *Spine J* 2013;13(8):889–93.
33. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KD. Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? *Clin Orthop Relat Res* 2010;468:57.
34. Baker PN, van der Meulen JH, Lewsey J, Gregg PJ. The role of pain and function in determining patient satisfaction after total knee replacement. Data from the National Joint Registry for England and Wales. *J Bone Joint Surg Br* 2007;89:893.
35. Nunez M, Lozano L, Nunez E, Segur JM, Sastre S, Macule F, et al. Total knee replacement and health-related quality of life: factors influencing long-term outcomes. *Arthritis Rheum* 2009;61:1062.